2.2: Structure & Function - Amino Acids

Source: BiochemFFA_2_1.pdf. The entire textbook is available for free from the authors at http://biochem.science.oregonstate.edu/content/biochemistry-free-and-easy

All of the proteins on the face of the earth are made up of the same 20 amino acids. Linked together in long chains called polypeptides, amino acids are the building blocks for the vast assortment of proteins found in all living cells.

"It is one of the more striking generalizations of biochemistry ...that the twenty amino acids and the four bases, are, with minor reservations, the same throughout Nature." - Francis Crick

All amino acids have the same basic structure, which is shown in Figure 2.1. At the “center” of each amino acid is a carbon called the α carbon and attached to it are four groups - a hydrogen, an α-carboxyl group, an α-amine group, and
an R-group, sometimes referred to as a side chain. The α carbon, carboxyl, and amino groups are common to all amino acids, so the R-group is the only unique feature in each amino acid. (A minor exception to this structure is that of proline, in which the end of the R-group is attached to the α-amine.) With the exception of glycine, which has an R-group consisting of a hydrogen atom, all of the amino acids in proteins have four different groups attached to them and consequently can exist in two mirror image forms, L and D. With only very minor exceptions, every amino acid found in cells and in proteins is in the L configuration.

Figure 2.1 - General amino acid structure

There are 22 amino acids that are found in proteins and of these, only 20 are specified by the universal genetic code. The others, selenocysteine and pyrrolysine use tRNAs that are able to base pair with stop codons in the mRNA during translation. When this happens, these unusual amino acids can be incorporated into proteins. Enzymes containing selenocysteine, for example, include glutathione peroxidases, tetraiodothyronine 5’ deiodinases, thioredoxin reductases, formate dehydrogenases, glycine reductases, and selenophosphate synthetase. Pyrrolysine-containing proteins are much rarer and are mostly confined to archaea.

Essential and non-essential

Nutritionists divide amino acids into two groups - essential amino acids (must be in the diet because cells can’t synthesize them) and non-essential amino acids (can be made by cells). This classification of amino acids has little to do with the structure of amino acids. Essential amino acids vary considerable from one organism to another and even differ in humans, depending on whether they are adults or children. Table 2.1 shows essential and non-essential amino acids in humans.

Some amino acids that are normally nonessential, may need to be obtained from the diet in certain cases. Individuals who do not synthesize sufficient amounts of arginine, cysteine, glutamine, proline, selenocysteine, serine, and tyrosine, due to illness, for example, may need dietary supplements containing these amino acids.
Non-protein amino acids

There are also α-amino acids found in cells that are not incorporated into proteins. Common ones include ornithine and citrulline. Both of these compounds are intermediates in the urea cycle. Ornithine is a metabolic precursor of arginine and citrulline can be produced by the breakdown of arginine. The latter reaction produces nitric oxide, an important signaling molecule. Citrulline is the metabolic byproduct. It is sometimes used as a dietary supplement to reduce muscle fatigue.

R-group chemistry

We separate the amino acids into categories based on the chemistry of their R-groups. If you compare groupings of amino acids in different textbooks, you will see different names for the categories and (sometimes) the same amino acid

Table 2.1 - Essential and non-essential amino acids

<table>
<thead>
<tr>
<th>Essential</th>
<th>Non-Essential</th>
</tr>
</thead>
<tbody>
<tr>
<td>Histidine</td>
<td>Alanine</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>Arginine</td>
</tr>
<tr>
<td>Leucine</td>
<td>Asparagine</td>
</tr>
<tr>
<td>Lysine</td>
<td>Aspartic acid</td>
</tr>
<tr>
<td>Methionine</td>
<td>Cysteine</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>Glutamic acid</td>
</tr>
<tr>
<td>Threonine</td>
<td>Glutamine</td>
</tr>
<tr>
<td>Tryptophan</td>
<td>Glycine</td>
</tr>
<tr>
<td>Valine</td>
<td>Proline</td>
</tr>
<tr>
<td></td>
<td>Selenocysteine</td>
</tr>
<tr>
<td></td>
<td>Serine</td>
</tr>
<tr>
<td></td>
<td>Tyrosine</td>
</tr>
</tbody>
</table>

Table 2.2 - Amino acid categories (based on R-group properties)
being categorized differently by different authors. Indeed, we categorize tyrosine both as an aromatic amino acid and as a hydroxyl amino acid. It is useful to classify amino acids based on their R-groups, because it is these side chains that give each amino acid its characteristic properties. Thus, amino acids with (chemically) similar side groups can be expected to function in similar ways, for example, during protein folding.

**Non-polar amino acids**

- **Alanine** (Ala/A) is one of the most abundant amino acids found in proteins, ranking second only to leucine in occurrence. A D-form of the amino acid is also found in bacterial cell walls. Alanine is non-essential, being readily synthesized from pyruvate. It is coded for by GCC, GCA, and GCG.

- **Glycine** (Gly/G) is the amino acid with the shortest side chain, having an R-group consistent only of a single hydrogen. As a result, glycine is the only amino acid that is not chiral. Its small side chain allows it to readily fit into both hydrophobic and hydrophilic environments.

![Figure 2.2 - Amino acid side chain properties Wikipedia](https://bio.libretexts.org/Bookshelves/Biochemistry/Book%3A_Biochemistry_Free_For_All_(Ahern_Rajagopal_and_Tan)/02%…) Updated: Sun, 01 Nov 2020 03:51:45 GMT

- Glycine is specified in the genetic code by GGU, GCC, GGA, and GGG. It is nonessential to humans.

- **Isoleucine** (Ile/I) is an essential amino acid encoded by AUU, AUC, and AUA. It has a hydrophobic side chain and is also chiral in its side chain.

- **Leucine** (Leu/L) is a branched-chain amino acid that is hydrophobic and essential. Leucine is the only dietary amino acid reported to directly stimulate protein synthesis in muscle, but caution is in order, as 1) there are conflicting studies and 2) leucine toxicity is dangerous, resulting in "the four D's": diarrhea, dermatitis, dementia and death. Leucine is encoded by six codons: UUA, UUG, CUU, CUC, CUA, CUG.

- **Methionine** (Met/M) is an essential amino acid that is one of two sulfur-containing amino acids - cysteine is the other. Methionine is non-polar and encoded solely by the AUG codon. It is the “initiator” amino acid in protein synthesis, being the first one incorporated into protein chains. In prokaryotic cells, the first methionine in a protein is formylated.

- Methionine is specified in the genetic code by UUA, UUG, CUU, CUC, CUA, CUG.
Proline (Pro/P) is the only amino acid found in proteins with an R-group that joins with its own α-amino group, making a secondary amine and a ring. Proline is a non-essential amino acid and is coded by CCU, CCC, CCA, and CCG. It is the least flexible of the protein amino acids and thus gives conformational rigidity when present in a protein. Proline’s presence in a protein affects its secondary structure. It is a disrupter of α-helices and β-strands. Proline is often hydroxylated in collagen (the reaction requires Vitamin C - ascorbate) and this has the effect of increasing the protein’s conformational stability. Proline hydroxylation of hypoxia-inducible factor (HIF) serves as a sensor of oxygen levels and targets HIF for destruction when oxygen is plentiful.

Valine (Val/V) is an essential, non-polar amino acid synthesized in plants. It is noteworthy in hemoglobin, for when it replaces glutamic acid at position number six, it causes hemoglobin to aggregate abnormally under low oxygen conditions, resulting in sickle cell disease. Valine is coded in the genetic code by GUU, GUC, GUA, and GUG.

Carboxyl Amino Acids

Aspartic acid (Asp/D) is a non-essential amino acid with a carboxyl group in its R-group. It is readily produced by transamination of oxaloacetate. With a pKa of 3.9, aspartic acid’s side chain is negatively charged at physiological pH. Aspartic acid is specified in the genetic code by the codons GAU and GAC.

Glutamic acid (Glu/E), which is coded by GAA and GAG, is a non-essential amino acid readily made by transamination of α-ketoglutarate. It is a neurotransmitter and has an R-group with a carboxyl group that readily ionizes (pKa = 4.1) at physiological pH.

Amine amino acids

Arginine (Arg/R) is an amino acid that is, in some cases, essential, but non-essential in others. Premature infants
cannot synthesize arginine. In addition, surgical trauma, sepsis, and burns increase demand for arginine. Most people, however, do not need arginine supplements. Arginine’s side chain contains a complex guanidinium group with a pKa of over 12, making it positively charged at cellular pH. It is coded for by six codons - CGU, CGC, CGA, CGG, AGA, and AGG.

- Histidine (His/H) is the only one of the proteinaceous amino acids to contain an imidazole functional group. It is an essential amino acid in humans and other mammals. With a side chain pKa of 6, it can easily have its charge changed by a slight change in pH. Protonation of the ring results in two NH structures which can be drawn as two equally important resonant structures.

![Figure 2.5 - Amine amino acids](https://bio.libretexts.org/Bookshelves/Biochemistry/Book%3A_Biochemistry_Free_For_All_(Ahern_Rajagopal_and_Tan)/02%…)

- Lysine (Lys/K) is an essential amino acid encoded by AAA and AAG. It has an Rgroup that can readily ionize with a charge of +1 at physiological pH and can be posttranslationally modified to form acetylysine, hydroxyllysine, and methyllysine. It can also be ubiquitinated, sumoylated, neddylated, biotinylated, carboxylated, and pupylated, and O-Glycosylation of hydroxyllysine is used to flag proteins for export from the cell. Lysine is often added to animal feed because it is a limiting amino acid and is necessary for optimizing growth of pigs and chickens.

### Aromatic amino acids

![Figure 2.6 - Aromatic amino acids](https://bio.libretexts.org/Bookshelves/Biochemistry/Book%3A_Biochemistry_Free_For_All_(Ahern_Rajagopal_and_Tan)/02%…)

- Phenylalanine (Phe/ F) is a non-polar, essential amino acid coded by UUU and UUC. It is a metabolic precursor
of tyrosine. Inability to metabolize phenylalanine arises from the genetic disorder known as phenylketonuria. Phenylalanine is a component of the aspartame artificial sweetener.

- Tryptophan (Trp/W) is an essential amino acid containing an indole functional group. It is a metabolic precursor of serotonin, niacin, and (in plants) the auxin phytohormone. Though reputed to serve as a sleep aid, there are no clear research results indicating this.
- Tyrosine (Tyr/Y) is a non-essential amino acid coded by UAC and UAU. It is a target for phosphorylation in proteins by tyrosine protein kinases and plays a role in signaling processes. In dopaminergic cells of the brain, tyrosine hydroxylase converts tyrosine to l-dopa, an immediate precursor of dopamine. Dopamine, in turn, is a precursor of norepinephrine and epinephrine. Tyrosine is also a precursor of thyroid hormones and melanin.

### Hydroxyl amino acids

- Serine (Ser/S) is one of three amino acids having an R-group with a hydroxyl in it (threonine and tyrosine are the others). It is coded by UCU, UCC, UCA, UGC, AGU, and AGC. Being able to hydrogen bond with water, it is classified as a polar amino acid. It is not essential for humans. Serine is precursor of many important cellular compounds, including purines, pyrimidines, sphingolipids, folate, and of the amino acids glycine, cysteine, and tryptophan. The hydroxyl group of serine in proteins is a target for phosphorylation by certain protein kinases. Serine is also a part of the catalytic triad of serine proteases.

- Threonine (Thr/T) is a polar amino acid that is essential. It is one of three amino acids bearing a hydroxyl group (serine and tyrosine are the others) and, as such, is a target for phosphorylation in proteins. It is also a target for Oglycosylation of proteins. Threonine proteases use the hydroxyl group of the amino acid in their catalysis and it is a precursor in one biosynthetic pathway for making glycine. In some applications, it is used as a pro-drug to increase brain glycine levels. Threonine is encoded in the genetic code by ACU, ACC, ACA, and ACG.

Tyrosine - see HERE.
Other amino acids

- Asparagine (Asn/N) is a non-essential amino acid coded by AAU and AAC. Its carboxyamide in the R-group gives it polarity. Asparagine is implicated in formation of acrylamide in foods cooked at high temperatures (deep frying) when it reacts with carbonyl groups. Asparagine can be made in the body from aspartate by an amidation reaction with an amine from glutamine. Breakdown of asparagine produces malate, which can be oxidized in the citric acid cycle.

- Cysteine (Cys/C) is the only amino acid with a sulfhydryl group in its side chain. It is nonessential for most humans, but may be essential in infants, the elderly and individuals who suffer from certain metabolic diseases. Cysteine’s sulfhydryl group is readily oxidized to a disulfide when reacted with another one. In addition to being found in proteins, cysteine is also a component of the tripeptide, glutathione. Cysteine is specified by the codons UGU and UGC.
• Glutamine (Gln/Q) is an amino acid that is not normally essential in humans, but may be in individuals undergoing intensive athletic training or with gastrointestinal disorders. It has a carboxyamide side chain which does not normally ionize under physiological pHs, but which gives polarity to the side chain. Glutamine is coded for by CAA and CAG and is readily made by amidation of glutamate. Glutamine is the most abundant amino acid in circulating blood and is one of only a few amino acids that can cross the blood-brain barrier.

• Selenocysteine (Sec/U) is a component of selenoproteins found in all kingdoms of life. It is a component in several enzymes, including glutathione peroxidases and thioredoxin reductases. Selenocysteine is incorporated into proteins in an unusual scheme involving the stop codon UGA. Cells grown in the absence of selenium terminate protein synthesis at UGAs. However, when selenium is present, certain mRNAs which contain a selenocysteine insertion sequence (SECIS), insert selenocysteine when UGA is encountered. The SECIS element has characteristic nucleotide sequences and secondary structure base-pairing patterns. Twenty five human proteins contain selenocysteine.

• Pyrrolysine (Pyl/O) is a twenty second amino acid, but is rarely found in proteins. Like selenocysteine, it is not coded for in the genetic code and must be incorporated by unusual means. This occurs at UAG stop codons. Pyrrolysine is found in methanogenic archaean organisms and at least one methane-producing bacterium. Pyrrolysine is a component of methane-producing enzymes.

---

**Ionizing groups**

pKa values for amino acid side chains are very dependent upon the chemical environment in which they are present. For example, the R-group carboxyl found in aspartic acid has a pKa value of 3.9 when free in solution, but can be as high as 14 when in certain environments inside of proteins, though that is unusual and extreme. Each amino acid has at least one ionizable amine group (α-amine) and one ionizable carboxyl group (α-carboxyl). When these are bound in a peptide bond, they no longer ionize. Some, but not all amino acids have R-groups that can ionize. The charge of a protein then arises from the charges of the α-amine group, the α-carboxyl group, and the sum of the charges of the ionized R-groups. Titration/ionization of aspartic acid is depicted in Figure 2.10. Ionization (or deionization) within a protein's structure can have significant effect on the overall conformation of the protein and, since structure is related to function, a major impact on the activity of a protein.
Most proteins have relatively narrow ranges of optimal activity that typically correspond to the environments in which they are found (Figure 2.11). It is worth noting that formation of peptide bonds between amino acids removes ionizable hydrogens from both the α- amine and α- carboxyl groups of amino acids. Thus, ionization/ deionization in a protein arises only from 1) the amino terminus; 2) carboxyl terminus; 3) R-groups; or 4) other functional groups (such as sulfates or phosphates) added to amino acids post-translationally - see below.

Carnitine

Not all amino acids in a cell are found in proteins. The most common examples include ornithine (arginine metabolism), citrulline (urea cycle), and carnitine (Figure 2.12). When fatty acids destined for oxidation are moved into the mitochondrion for that purpose, they travel across the inner membrane attached to carnitine. Of the two stereoisomeric forms, the L form is the active one. The molecule is synthesized in the liver from lysine and methionine.

From exogenous sources, fatty acids must be activated upon entry into the cytoplasm by being joined to coenzyme A. The CoA portion of the molecule is replaced by carnitine in the intermembrane space of the mitochondrion in a reaction catalyzed by carnitine acyltransferase I. The resulting acylcarnitine molecule is transferred across the inner mitochondrial membrane by the carnitineacylcarnitine translocase and then in the matrix of the mitochondrion, carnitine acyltransferase II replaces the carnitine with coenzyme A (Figure 6.88).
Catabolism of amino acids

We categorize amino acids as essential or non-essential based on whether or not an organism can synthesize them. All of the amino acids, however, can be broken down by all organisms. They are, in fact, a source of energy for cells, particularly during times of starvation or for people on diets containing very low amounts of carbohydrate. From a perspective of breakdown (catabolism), amino acids are categorized as glucogenic if they produce intermediates that can be made into glucose or ketogenic if their intermediates are made into acetyl-CoA. Figure 2.13 shows the metabolic fates of catabolism of each of the amino acids. Note that some amino acids are both glucogenic and ketogenic.

![Figure 2.13 - Catabolism of amino acids. Some have more than one path. Image by Pehr Jacobson](https://bio.libretexts.org/Bookshelves/Biochemistry/Book%3A_Biochemistry_Free_For_All_(Ahern_Rajagopal_and_Tan)/02%…)

Post-translational modifications

After a protein is synthesized, amino acid side chains within it can be chemically modified, giving rise to more diversity of structure and function (Figure 2.14). Common alterations include phosphorylation of hydroxyl groups of serine, threonine, or tyrosine. Lysine, proline, and histidine can have hydroxyls added to amines in their R-groups. Other modifications to amino acids in proteins include addition of fatty acids (myristic acid or palmitic acid), isoprenoid groups, acetyl groups, methyl groups, iodine, carboxyl groups, or sulfates. These can have the effects of ionization (addition of phosphates/sulfates), deionization (addition of acetyl group to the R-group amine of lysine), or have no effect on charge at all. In addition, N-linked- and O-linked-glycoproteins have carbohydrates covalently attached to side chains of asparagine and threonine or serine, respectively.

Some amino acids are precursors of important compounds in the body. Examples include epinephrine, thyroid hormones, Ldopa, and dopamine (all from tyrosine), serotonin (from tryptophan), and histamine (from histidine).
Building Polypeptides

Although amino acids serve other functions in cells, their most important role is as constituents of proteins. Proteins, as we noted earlier, are polymers of amino acids.

Amino acids are linked to each other by peptide bonds, in which the carboxyl group of one amino acid is joined to the amino group of the next, with the loss of a molecule of water. Additional amino acids are added in the same way, by formation of peptide bonds between the free carboxyl on the end of the growing chain and the amino group of the next amino acid in the sequence. A chain made up of just a few amino acids linked together is called an oligopeptide (oligo=few) while a typical protein, which is made up of many amino acids is called a polypeptide (poly=many). The end of the peptide that has a free amino group is called the N-terminus (for NH2), while the end with the free carboxyl is termed the C-terminus (for carboxyl).
As we’ve noted before, function is dependent on structure, and the string of amino acids must fold into a specific 3-D shape, or conformation, in order to make a functional protein. The folding of polypeptides into their functional forms is the topic of the next section.